

日本国特許庁
JAPAN PATENT OFFICE

CERTIFIED COPY OF
PRIORITY DOCUMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office.

出願年月日
Date of Application:

2002年11月29日

出願番号
Application Number:

特願2002-346977

[ST. 10/C]:

[JP2002-346977]

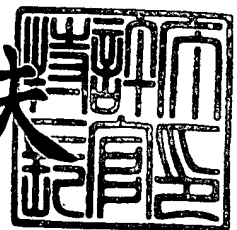
出願人
Applicant(s):


沖電気工業株式会社

2003年11月28日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫





【書類名】 特許願

【整理番号】 MA001413

【提出日】 平成14年11月29日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/20

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 佐々木 美樹

【発明者】

【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

【氏名】 池野 篤司

【特許出願人】

【識別番号】 000000295

【氏名又は名称】 沖電気工業株式会社

【代表者】 篠塚 勝正

【代理人】

【識別番号】 100083840

【弁理士】

【氏名又は名称】 前田 実

【選任した代理人】

【識別番号】 100116964

【弁理士】

【氏名又は名称】 山形 洋一

【手数料の表示】

【予納台帳番号】 007205

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9003703

【包括委任状番号】 0101807

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 数値情報検索装置

【特許請求の範囲】

【請求項 1】 自然言語で数値情報を検索する数値情報検索装置において、
検索対象の文書や検索したい数量表現を入力する入力手段と、
属性を表す属性名と属性を付与するための条件とが格納された属性情報辞書と

、
上記入力された文書や数値表現に上記属性情報辞書を参照して属性を付与する
属性付与手段と、

属性を表す属性名と変換前文字列と変換後文字列とが格納された変換情報辞書
と、

上記属性付与後の文書や数値表現に対して、上記変換辞書情報を参照して、属
性付与部分の数量を変換前文字列から変換後文字列に変換することによって、属
性付与部分の数値情報を変換する数量変換手段と

を備えた

ことを特徴とする数値情報検索装置。

【請求項 2】 請求項 1 記載の数値情報検索装置において、

上記属性付与変換後の文書から特定の属性部分を検索キーワードとして抽出す
る抽出手段と、

文書データが格納される文書データベースと、

上記属性付与変換後の文書と上記入力された元の文書と上記抽出された検索キ
ーワードとを上記文書データベースに格納する文書格納検索手段と

をさらに備えた

ことを特徴とする数値情報検索装置。

【請求項 3】 請求項 2 記載の数値情報検索装置において、

検索結果を出力する出力手段をさらに備え、

上記文書格納検索手段は、上記属性付与変換後の数値表現と検索キーワードが
一致する文書を上記文書データベースから検索し、上記検索結果として上記出力
手段に送る

ことを特徴とする数値情報検索装置。

【請求項 4】 請求項 2 記載の数値情報検索装置において、

属性を表す属性名と範囲の表現を表す範囲指定語と比較の大小関係を表す比較条件と出力の並べ替えのための順序を表す整列順とが格納された抽出情報辞書と

出力の並べ替えのための並べ替え手段と

検索結果を出力する出力手段と

をさらに備え、

上記抽出手段は、上記属性付与変換後の数値表現から属性名と数値と範囲指定語とを抽出し、

上記並べ替え手段は、上記数値表現から抽出された数値を、上記数値表現から抽出された属性名と属性名が一致する複数の検索キーワードの数値とそれぞれ比較し、上記抽出情報辞書を参照して、上記比較結果が範囲指定語の比較条件に適合する検索キーワードを選択し、選択した検索キーワードを整列順に従って並べ替え、

上記文書格納検索手段は、上記数値表現から抽出された属性名と属性名が一致する検索キーワードを上記文書データベースから検索して上記並べ替え手段に送り、上記並べ替えされた検索キーワードと一致する文書を上記文書データベースから検索し、これらの検索した文書を上記並べ替えされた検索キーワードの順序に従って上記検索結果として上記出力手段に送る

ことを特徴とする数値情報検索装置。

【請求項 5】 請求項 1 から 4 までのいずれかに記載の数値情報検索装置において、

上記属性付与手段は、上記属性情報辞書を参照して、属性の開きタグと閉じタグを原文の条件部分の前後に付与することを特徴とする数値情報検索装置。

【請求項 6】 請求項 1 から 4 までのいずれかに記載の数値情報検索装置において、

上記属性付与手段は、上記属性情報辞書を参照して、原文を変更することなく、原文中の条件部分の属性を示す属性情報データを上記原文とは別に保持できる

ように付与することを特徴とする数値情報検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、自然言語で数値情報を検索する数量表現検索装置に関する。

【0002】

【従来の技術】

自然言語で数値情報を検索できるようにするには、数量的な表現に対応しなければならない。

【0003】

従来の数値情報検索装置では、文脈情報と時刻／数量表現の基準と範囲に関する情報に着目することによって、相対的な表現を含んだ数量／時間表現を入力文として扱うことができる（例えば、特許文献1参照）。このような従来の数値情報検索装置では、例えば、「1990年」に対して「翌年」とは「1991年」であると補完され、「1位から3位まで」に対して「次」とは「4位から」であると補完される。

【0004】

【特許文献1】

特開平7-200583号公報

【0005】

【発明が解決しようとする課題】

しかしながら、上記従来の数値情報検索装置では、単位のように表記や基準が多様な数値的な表記や基準が多様な数値的な表現には対応することができない。

【0006】

例えば、「100円増加」は、「1000円」に対して「100円増加」であれば「+100」でよいが、「1メートル」に対して「1センチメートル増加」や「1mm増加」であれば、まず数値的な変換をしなければ補完することができない。

【0007】

本発明は、このような従来の従来の課題に鑑みてなされたものであり、表記や単位系を意識しなくても数値情報を検索することができる数値情報検索装置を提供することを目的とするものである。

【0008】

【課題を解決するための手段】

本発明の数値情報検索装置は、自然言語で数値情報を検索する数値情報検索装置において、

検索対象の文書や検索したい数量表現を入力する入力手段と、

属性を表す属性名と属性を付与するための条件とが格納された属性情報辞書と

、
上記入力された文書や数値表現に上記属性情報辞書を参照して属性を付与する属性付与手段と、

属性を表す属性名と変換前文字列と変換後文字列とが格納された変換情報辞書と、

上記属性付与後の文書や数値表現に対して、上記変換辞書情報を参照して、属性付与部分の数量を変換前文字列から変換後文字列に変換することによって、属性付与部分の数値情報を変換する数量変換手段と

を備えた

ことを特徴とするものである。

【0009】

【発明の実施の形態】

実施の形態 1

図1は本発明の実施の形態1の数値情報検索装置のブロック構成図である。この実施の形態1の数値情報検索装置は、入力手段1と、属性付与手段2と、属性情報辞書3と、数量変換手段4と、変換情報辞書5と、文書格納検索手段6と、文書データベース7と、抽出手段8と、出力手段9とを備える。

【0010】

入力手段1は、検索対象の文書や検索したい数値表現を入力する手段である。この入力手段1は、入力された文書や数値表現を属性付与手段2に送る。

【0011】

属性付与手段2は、属性を付与する手段である。この属性付与手段2は、属性情報辞書3を参照して、入力手段1から入力された文書や数値表現の条件部分の前後に属性の開きタグと閉じタグを付与して、数量変換手段4に送る。

【0012】

属性情報辞書3は、属性の情報を格納する辞書である。この属性情報辞書3は、属性情報は、属性名と条件からなる。

【0013】

数量変換手段4は、数値情報を変換する手段である。この数量変換手段4は、変換情報辞書5を参照して、属性付与手段2から入力された属性付与後の文書や数値表現に対して、属性の変換前文字列の部分を変換後文字列の部分に変換して、抽出手段8に送る。

【0014】

変換情報辞書5は、数量の変換の情報を格納する辞書である。この変換情報辞書5、属性名と変換前文字列と変換後文字列からなる。

【0015】

文書格納検索手段6は、文書を格納および検索する手段である。この文書格納検索手段6は、抽出手段8から入力された属性付与変換後の文書や元の文書や検索キーワードを文書データベース7に格納し、抽出手段8から入力された属性付与変換後の数値表現が検索キーワードと一致する文書を文書データベース7から検索して出力手段9に送る。

【0016】

文書データベース7は、検索対象の文書や属性付与変換後の文書を格納するデータベースである。

【0017】

抽出手段8は、キーワードや該当部分を抽出する手段である。この抽出手段8は、数量変換手段4から入力された変換後の文書や数値表現と、変換後の属性の部分を抽出した検索キーワードを、文書格納検索手段6に送る。

【0018】

出力手段 9 は、結果を出力する手段である。この出力手段 9 は、文書格納検索手段 6 から送られた検索結果を出力する。

【0019】

なお、属性を付与する処理、文書をデータベース化する処理、文書を格納および検索する処理、ならびに該当部分を抽出する処理は、一般的な部分に関しては、既知の自然言語処理技術を用いればよい。

【0020】

図 2 は本発明の実施の形態 1 の数値情報検索装置の動作を説明するフローチャートである。この図 2 において、入力手段 1 で処理を選択し（ステップ 201）、投入処理を実行するか（ステップ 202）、検索処理を実行するか（ステップ 203）、または終了する。

【0021】

図 3 は図 2 のステップ 202 の投入処理の動作を説明するフローチャートであり、図 4 は図 2 のステップ 203 の検索処理の動作を説明するフローチャートである。また、図 5 は属性情報辞書 3 の構成例を示す図であり、図 6 は変換情報辞書 5 の構成例を示す図である。なお、図 6 の変換情報辞書 5 の変換後文字列においての「*10** - 9 m」, 「*10** - 3 m」は、それぞれ「×10 - 9 m」, 「×10 - 3 m」を示す。

【0022】

図 3 の投入処理では、まず最初に、検索対象の文書を入力手段 1 に投入する（ステップ 301）。

例えば、以下の例文（a）、（b）、または（c）を投入する。

「A社は80ナノメートルの微細加工技術の開発に成功」…（a）

「B社は100nmが限界」…（b）

「0.1mmが限界だった時代は遠く」…（c）

入力手段 1 に投入された文書は、属性付与手段 2 に送られる。

【0023】

次に、属性付与手段 2 において、属性情報辞書 3（図 5 参照）を参照して、投入された文書から条件部分を探し、その条件部分に属性を付与する（ステップ 3

0 2)。

例文 (a) では、条件部分として「8 0 ナノメートル」が含まれており、その属性名は「LENGTH」なので、その条件部分の前後に属性の開きタグ<LENGTH>および属性の閉じタグ</LENGTH>をそれぞれ付与し、

「A社は<LENGTH>8 0 ナノメートル</LENGTH>の微細加工技術の開発に成功」

となる。

例文 (b) では、条件部分として「1 0 0 n m」が含まれており、その属性名は「LENGTH」なので、その条件部分の前後にその属性のタグを付与し、

「B社は<LENGTH>1 0 0 n m</LENGTH>が限界」

となる。

例文 (c) では、条件部分として「0. 1 m m」が含まれているので、同様に

「<LENGTH>0. 1 m m</LENGTH>が限界だった時代は遠く」

となる。

属性付与手段 2 で属性付与された文書は、数量変換手段 4 に送られる。

【0 0 2 4】

次に、数量変換手段 4 において、変換情報辞書 5 (図 6 参照) を参照して、文書の属性の数量を変換する (ステップ 3 0 3)。

例文 (a) では、属性の数量 (変換前文字列) 「ナノメートル」が変換後文字列「 $\times 10^{-9}$ m」に変換され、

「A社は<LENGTH>8 0 $\times 10^{-9}$ m</LENGTH>の微細加工技術の開発に成功」

となる。

例文 (b) では、変換前文字列「n m」が変換後文字列「 $\times 10^{-9}$ m」に変換され、

「B社は<LENGTH>1 0 0 $\times 10^{-9}$ m</LENGTH>が限界」

となる。

例文 (c) では、変換前文字列「m m」が変換後文字列「 $\times 10^{-3}$ m」に変

換され、

「 $\langle \text{LENGTH} \rangle 0.1 \times 10^{-3} \text{ m} \langle / \text{LENGTH} \rangle$ が限界だった時代は遠く」

となる。

数量変換手段4で数量変換された文書は、抽出手段8に送られる。

【0025】

次に、抽出手段8において、数量変換後の属性の部分を検索キーワードとして抽出する（ステップ304）。

例文（a）では、「 $\langle \text{LENGTH} \rangle 80 \times 10^{-9} \text{ m} \langle / \text{LENGTH} \rangle$ 」が検索キーワードとしてが抽出される。

例文（b）では、「 $\langle \text{LENGTH} \rangle 100 \times 10^{-9} \text{ m} \langle / \text{LENGTH} \rangle$ 」が検索キーワードとしてが抽出される。

例文（c）では、「 $\langle \text{LENGTH} \rangle 0.1 \times 10^{-3} \text{ m} \langle / \text{LENGTH} \rangle$ 」が検索キーワードとしてが抽出される。

抽出手段8で抽出された検索キーワードは、元の文書および属性付与変換後の文書とともに文書格納検索手段6に送られる。

【0026】

最後に、文書格納検索手段6によって、文書データベース7に、元の文書と属性付与変換後の文書と検索キーワードを格納し（ステップ305）、投入処理を終了する。

例文（a）では、元の文書「A社は80ナノメートルの微細加工技術の開発に成功」と、属性付与変換後の文書「A社は $\langle \text{LENGTH} \rangle 80 \times 10^{-9} \text{ m} \langle / \text{LENGTH} \rangle$ の微細加工技術の開発に成功」と、検索キーワード「 $\langle \text{LENGTH} \rangle 80 \times 10^{-9} \text{ m} \langle / \text{LENGTH} \rangle$ 」とが、文書データベース7に格納される。

例文（b）では、元の文書「B社は100nmが限界」と、属性付与変換後の文書「B社は $\langle \text{LENGTH} \rangle 100 \times 10^{-9} \text{ m} \langle / \text{LENGTH} \rangle$ が限界」と、検索キーワード「 $\langle \text{LENGTH} \rangle 100 \times 10^{-9} \text{ m} \langle / \text{LENGTH} \rangle$ 」とが、文書データベース7に格納される。

例文（c）では、元の文書「0.1mmが限界だった時代は遠く」と、属性付与変換後の文書「 $\langle \text{LENGTH} \rangle 0.1 \times 10^{-3} \text{ m} \langle / \text{LENGTH} \rangle$ が限界だった時代は遠

く」と、検索キーワード「<LENGTH>0.1×10⁻³m</LENGTH>」とが、文書データベース7に格納される。

【0027】

図4の検索処理では、まず最初に、検索したい数値表現を、検索語として入力手段1に入力する（ステップ401）。

例えば、以下の例文（d）を検索語として入力する。

「80nm」…（d）

入力手段1に入力された数値表現（検索語）は、属性付与手段2に送られる。

【0028】

次に、属性付与手段2において、属性情報辞書3（図5参照）を参照して、検索語に属性を付与する（ステップ402）。

例文（d）では、その属性名は「LENGTH」なので、その条件部分の前後に属性の開きタグ<LENGTH>および属性の閉じタグ</LENGTH>をそれぞれ付与し、

「<LENGTH>80nm</LENGTH>」

となる。

属性付与手段2で属性付与された数値表現（検索語）は、数量変換手段4に送られる。

【0029】

次に、数量変換手段4において、変換情報辞書5（図6参照）を参照して、検索語の属性の数量を変換する（ステップ403）。

例文（d）では、属性の数量（変換前文字列）「nm」が変換後文字列「×10⁻⁹m」に変換され、

「<LENGTH>80×10⁻⁹m</LENGTH>」

となる。

数量変換手段4で数量変換された検索語は、抽出手段8を介して文書格納検索手段6に送られる。

【0030】

次に、文書格納検索手段6において、数量変換後の検索語と検索キーワードが一致する文書を、文書データベース7に格納されている文書から検索して獲得す

る（ステップ404）。

例文（d）では、検索キーワードが「<LENGTH>80×10⁻⁹m</LENGTH>」である上記例文（a）の文書「A社は80ナノメートルの微細加工技術の開発に成功」が、数量変換後の検索語「<LENGTH>80×10⁻⁹m</LENGTH>」と検索キーワードが一致する文書として、文書データベース7から検索されて獲得される。

【0031】

そして、上記ステップ404で獲得した文書を検索結果として出力手段9から出力する（ステップ405）。

【0032】

なお、出力するのは、属性の部分のみでも、属性の前後の適当な範囲でも、文書全体でもよいし、元の文書からでも、属性付与後の文書からでも、数量変換後の文書からでもよい。

【0033】

以上のように実施の形態1によれば、数値的な変換をすることによって、表記や単位系を意識しなくても数値情報を検索できる。

【0034】

実施の形態2

図7は本発明の実施の形態2の数値情報検索装置のブロック構成図である。なお、図7において、図1と同じものあるいは相当するものには同じ符号を付してある。この実施の形態2の数値情報検索装置は、入力手段1と、属性付与手段2と、属性情報辞書3と、数量変換手段4と、変換情報辞書5と、文書格納検索手段6と、文書データベース7と、抽出手段8と、出力手段9と、並べ替え手段10と、抽出情報辞書11とを備えており、上記実施の形態1の数値情報検索装置（図1参照）において、並べ替え手段10および抽出情報辞書11を設けたものである。

【0035】

並べ替え手段10は、抽出結果を並べ替える手段である。抽出情報辞書11は、抽出と並べ替えの情報を格納する辞書である。抽出情報辞書11は、属性名と

範囲指定語と比較条件と整列順からなる。並べ替え手段10は、抽出情報辞書11を参照して、文書格納検索手段6から送られた検索キーワードと属性の数値を比較して、範囲指定語の比較条件に該当する検索キーワードを抽出し、整列順に従って並べ替えて、文書データベース7からされた文書を出力手段9に送る。

【0036】

この実施の形態2の数値情報検索装置の全体の動作および投入処理の動作は、上記実施の形態1の全体の動作（図2参照）および投入処理の動作（図3参照）と同様である。実施の形態2の検索処理の動作について以下に説明する。

【0037】

図8は本発明の実施の形態2の数値情報検索装置における検索処理の動作を説明するフローチャートである。また、図9は抽出情報辞書11の構成例を示す図である。なお、図9の抽出情報辞書11の比較条件における「<=」，「=」は、それぞれ「≤」，「=」を示す。

【0038】

以下の実施の形態2の検索処理の説明においては、上記実施の形態1で説明した例文（a），（b），（c）の投入処理がすでになされているものとする。

【0039】

図8の検索処理では、まず最初に、検索したい数値表現を、検索語として入力手段1に入力する（ステップ801）。

例えば、以下の例文（e）を検索語として入力する。

「100nm以下」…（e）

入力手段1に入力された数値表現（検索語）は、属性付与手段2に送られる。

【0040】

次に、属性付与手段2において、属性情報辞書3（図5参照）を参照して、検索語から条件部分を探し、その条件部分に属性を付与する（ステップ802）。

例文（e）では、条件部分として「100nm」が含まれており、その属性名は「LENGTH」なので、その条件部分の前後に属性の開きタグ<LENGTH>および属性の閉じタグ</LENGTH>をそれぞれ付与し、

「<LENGTH>100nm</LENGTH>以下」

となる。

属性付与手段 2 で属性付与された数値表現（検索語）は、数量変換手段 4 に送られる。

【0041】

次に、数量変換手段 4 において、変換情報辞書 5（図 6 参照）を参照して、検索語の属性の数量を変換する（ステップ 403）。

例文（e）では、属性の数量（変換前文字列）「nm」が変換後文字列「 $\times 10^{-9}m$ 」に変換され、

「<LENGTH>100 $\times 10^{-9}$ </LENGTH>以下」

となる。

数量変換手段 4 で数量変換された検索語は、抽出手段 8 に送られる。

【0042】

次に、抽出手段 8 において、検索語の数量変換後の属性名と数値と範囲指定語を抽出する（ステップ 804）。

例文（e）では、属性名として「LENGTH」が抽出され、数値として「 $100 \times 10^{-9}m$ 」が抽出され、範囲指定語として「以下」が抽出される。

抽出手段 8 で抽出された属性名と数値と範囲指定語は、文書格納検索手段 6 に送られる。

【0043】

次に、文書格納検索手段 6 において、検索語から抽出された属性名と属性名が一致する検索キーワードを、文書データベース 7 に格納されている検索キーワードから検索する（ステップ 805）。

例文（e）では、属性名が「LENGTH」である検索キーワードの検索がなされ、上記例文（a）の検索キーワード「<LENGTH>80 $\times 10^{-9}m$ </LENGTH>」と、上記例文（b）の検索キーワード「<LENGTH>100 $\times 10^{-9}m$ </LENGTH>」と、上記例文（c）の検索キーワード「<LENGTH>0.1 $\times 10^{-3}m$ </LENGTH>」とが、文書データベース 7 から検索される。

文書データベース 7 から検索された検索キーワードは、属性付与変換後の検索語とともに並べ替え手段 10 に送られる。

【0044】

次に、並べ替え手段10において、抽出情報辞書（図9参照）を参照して、検索語の数値と文書データベース7から検索された検索キーワードの数値とを比較する（ステップ806）。

例文（e）では、検索語の数値は「 $100 \times 10^{-9} \text{ m}$ 」であって、

上記例文（a）の検索キーワードの数値「 $80 \times 10^{-9} \text{ m}$ 」との比較では、

$$80 \times 10^{-9} \text{ m} \text{ 「<」 } 100 \times 10^{-9} \text{ m}$$

であり、

上記例文（b）の検索キーワードの数値「 $100 \times 10^{-9} \text{ m}$ 」との比較では、

$$100 \times 10^{-9} \text{ m} \text{ 「=」 } 100 \times 10^{-9} \text{ m}$$

であり、

上記例文（c）の検索キーワードの数値「 $0.1 \times 10^{-3} \text{ m}$ 」との比較では、

$$0.1 \times 10^{-3} \text{ m} \text{ 「>」 } 100 \times 10^{-9} \text{ m}$$

である。

【0045】

次に、並べ替え手段10において、抽出情報辞書（図9参照）を参照して、比較した結果が範囲指定語の比較条件に適合する検索キーワードを選択する（ステップ807）。

例文（e）では、範囲指定語「以下」の比較条件に適合するには、比較結果が「<」である上記例文（a）の検索キーワード「<LENGTH> $80 \times 10^{-9} \text{ m}$ </LENGTH>」と、比較結果が「=」である上記例文（b）の検索キーワード「<LENGTH> $100 \times 10^{-9} \text{ m}$ </LENGTH>」であり、これら2つの検索キーワードが選択される。

【0046】

次に、並べ替え手段10において、抽出情報辞書（図9参照）を参照して、選択した検索キーワードを整列順に従って並べ替える（ステップ808）。

例文（e）では、その範囲指定語「以下」の整列順は「昇順」に設定されてい

るので（図9参照）、上記2つの検索キーワードは、

「<LENGTH>80×10⁻⁹m</LENGTH>」 （上記例文（a）の検索キーワード）

「<LENGTH>100×10⁻⁹m</LENGTH>」 （上記例文（b）の検索キーワード）

の順で並べ替えられる。

並べ替えられた検索キーワードは、文書格納検索手段6に送られる。

【0047】

次に、文書格納検索手段6において、並べ替えられた検索キーワードと検索キーワードが一致する文書を、文書データベース7に格納されている文書から検索して獲得する（ステップ809）。

【0048】

そして、上記ステップ809で獲得した文書を、並べ替えられた検索キーワードの順に従って、検索結果として出力手段9から出力する（ステップ810）。

上記例文（e）では、

「A社は80ナノメートルの微細加工技術の開発に成功」（上記例文（a））

「B社は100nmが限界」（上記例文（b））

の順で出力される。

【0049】

なお、以下の例文（f）を検索語として入力した場合には、範囲指定語は「かそれ以下」となる（図9参照）。

「100nmかそれ以下」…（f）

この例文（f）では、範囲指定語「かそれ以下」の整列順は「降順」に設定されているので（図9参照）、

「<LENGTH>100×10⁻⁹m</LENGTH>」（上記例文（b）の検索キーワード）

「<LENGTH>80×10⁻⁹m</LENGTH>」（上記例文（a）の検索キーワード）

の順に検索キーワードが並べ替えられて、

「B社は100nmが限界」（上記例文（b））

「A社は80ナノメートルの微細加工技術の開発に成功」（上記例文（a））
の順で出力され、範囲指定語が「以下」の場合とは逆となる。

【0050】

以上のように実施の形態2によれば、範囲の表現を数値の大小関係に変換して比較することによって、数式を意識しなくても数値情報を検索することができる。また、範囲の表現に従って検索結果を並べ替えることにより、効率よく検索結果を確認できる。また、自然言語の検索に用いることによって、数量表現の検索が容易になる。

【0051】

なお、上記実施の形態1および2では、属性の付与には属性タグを原文に挿入する方法を用いて説明したが、別の方法を用いてもよい。例えば、原文を変更することなく、「原文の35文字目から40文字目までが「LENGTH」の属性を持つ」といった情報が記載された原文の属性情報データとして、原文とは独立して保持する方法などが考えられる。

【0052】

また、上記実施の形態1および2では、数値情報のみを検索語および検索対象とする数値情報検索装置を説明したが、本発明は、数値以外の語を検索語または検索対象とする方法や装置と組み合わせて利用してもよい。

【0053】

【発明の効果】

以上のように本発明によれば、数値的な変換をすることによって、表記や単位系を意識しなくても数値情報を検索できるという効果がある。

【図面の簡単な説明】

【図1】 本発明の実施の形態1の数値情報検索装置のブロック構成図である。

【図2】 図1の数値情報検索装置の動作を説明するフローチャートである。

【図3】 図2のステップ202の投入処理の動作を説明するフローチャー

トである。

【図 4】 図 2 のステップ 203 の検索処理の動作を説明するフローチャートである。

【図 5】 図 1 の属性情報辞書の構成例を示す図である。

【図 6】 図 1 の変換情報辞書の構成例を示す図である。

【図 7】 本発明の実施の形態 2 の数値情報検索装置のブロック構成図である。

【図 8】 図 7 の数値情報検索装置においての検索処理の動作を説明するフローチャートである。

【図 9】 図 7 の抽出情報辞書の構成例を示す図である。

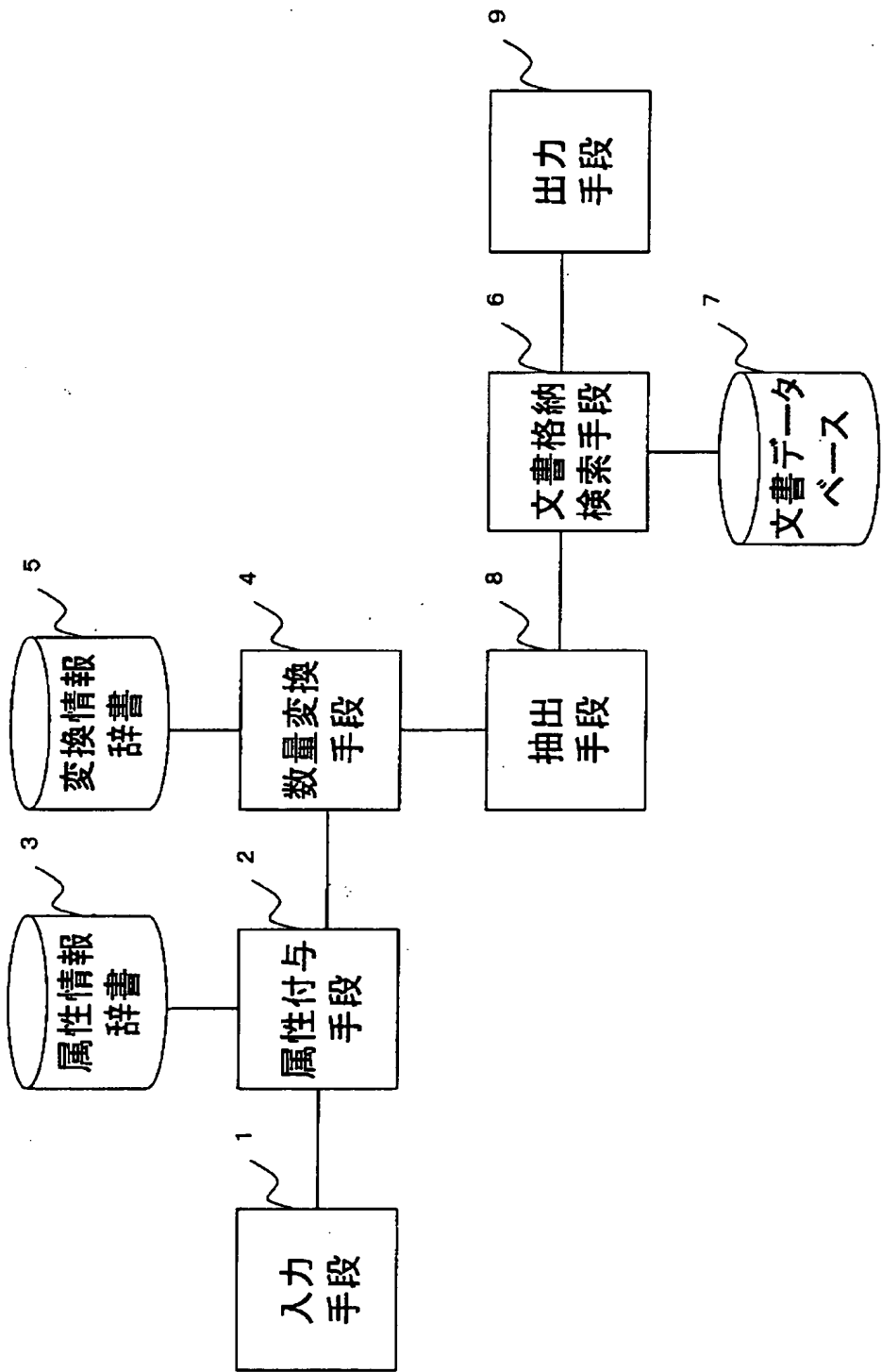
【符号の説明】

1 入力手段、 2 属性付与手段、 3 属性情報辞書、 4 数量変換手段、 5 変換情報辞書、 6 文書格納検索手段、 7 文書データベース、 8 抽出手段、 9 出力手段、 10 並べ替え手段、 11 抽出情報辞書。

【書類名】

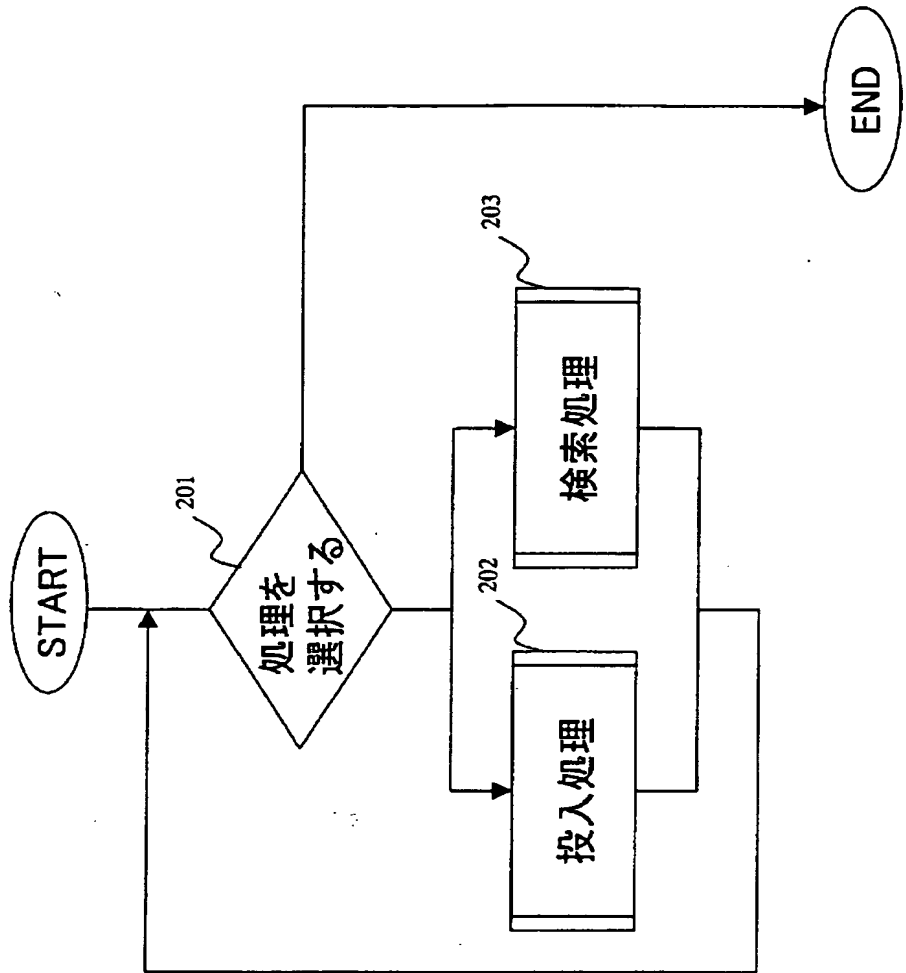
図面

【図 1】



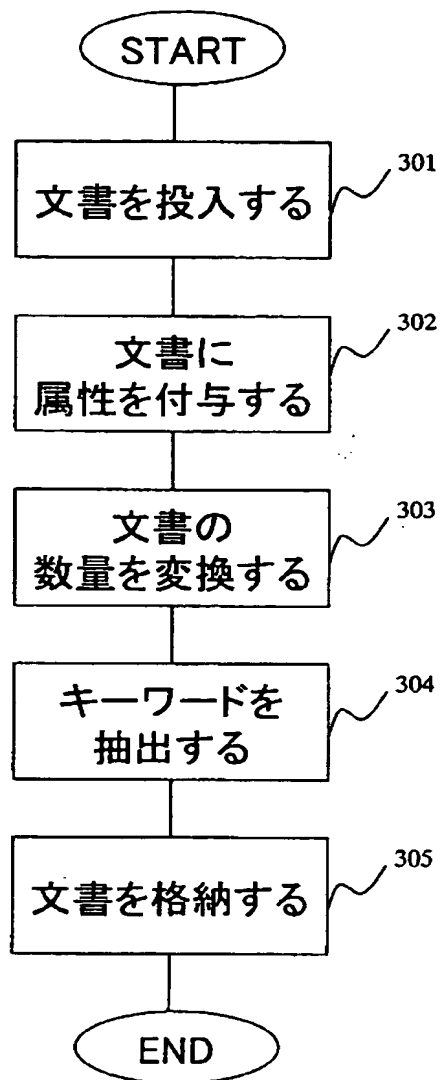
実施の形態 1 の構成

【図 2】



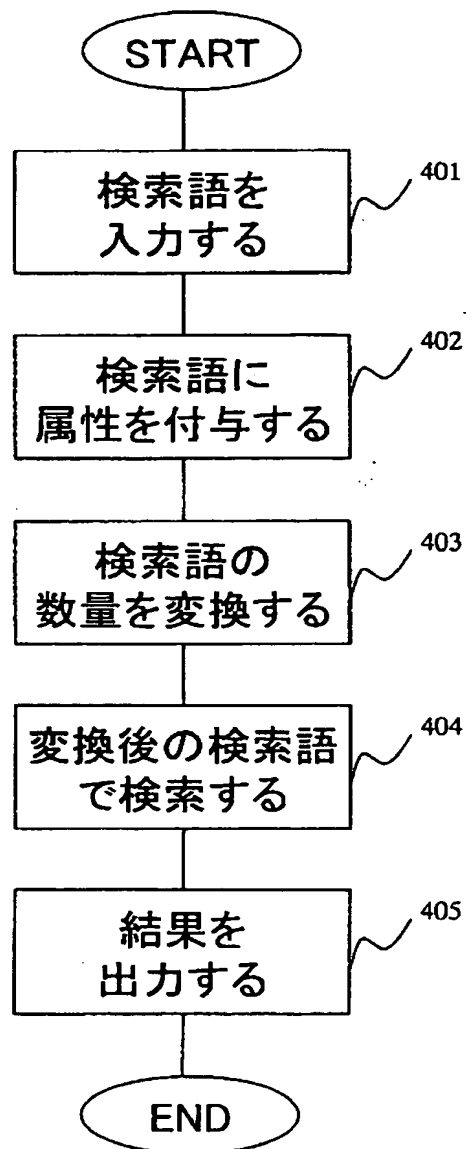
実施の形態 1 の全体の流れ

【図 3】



実施の形態 1 の投入処理の流れ

【図 4】



実施の形態 1 の検索処理の流れ

【図 5】

属性名	条件
LENGTH	数 + “ナノメートル”
LENGTH	数 + “nm”
LENGTH	数 + “mm”
...

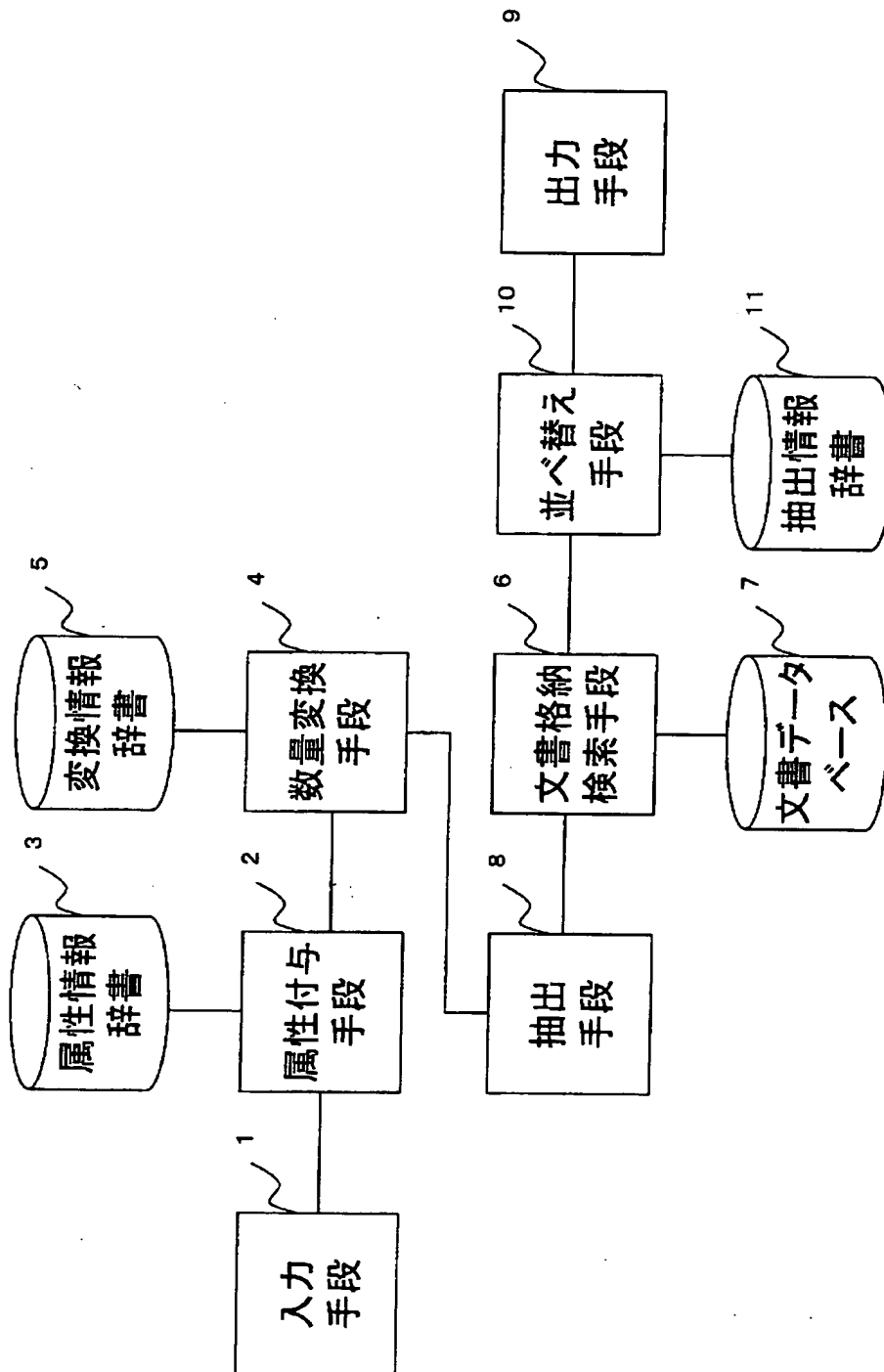
属性情報辞書の例

【図 6】

属性名	変換前文字列	変換後文字列
LENGTH	ナノメートル	*10**-9m
LENGTH	nm	*10**-9m
LENGTH	mm	*10**-3m
...

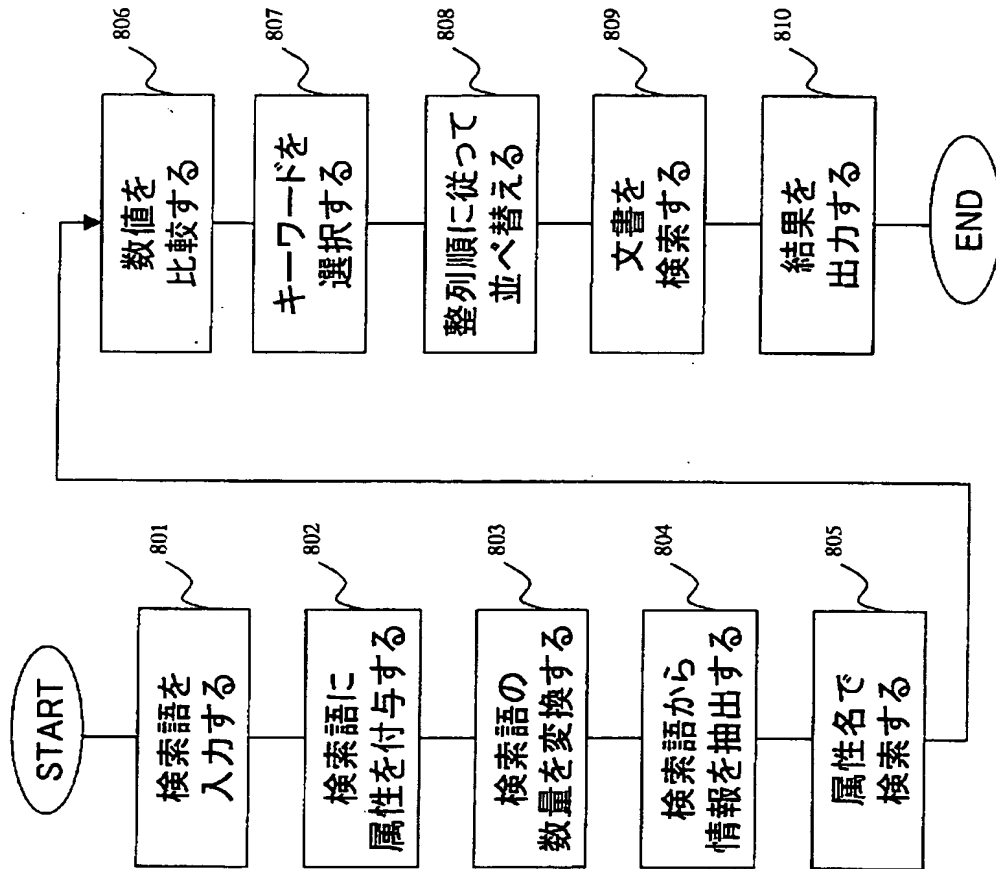
変換情報辞書の例

【図 7】



実施の形態 2 の構成

【図 8】



実施の形態 2 の検索処理の流れ

【図 9】

属性名	範囲指定語	比較条件	整列順
LENGTH	以下	<=、<、==	昇順
LENGTH	かそれ以下	<=、<、==	降順
LENGTH	未満	<	昇順
...

抽出情報辞書の例

【書類名】 要約書

【要約】

【課題】 表記や単位系を意識しなくても数値情報を検索することができる数値情報検索装置を提供する。

【解決手段】 検索対象の文書や検索したい数量表現を入力する入力手段 1 と、属性を表す属性名と属性を付与するための条件とが格納された属性情報辞書 3 と、上記入力された文書や数値表現に上記属性情報辞書を参照して属性を付与する属性付与手段 2 と、属性を表す属性名と変換前文字列と変換後文字列とが格納された変換情報辞書 5 と、上記属性付与後の文書や数値表現に対して、上記変換辞書情報を参照して、属性付与部分の数量を変換前文字列から変換後文字列に変換することによって、属性付与部分の数値情報を変換する数量変換手段 4 とを備える。

【選択図】 図 1

特願 2 0 0 2 - 3 4 6 9 7 7

出 願 人 履 歷 情 報

識別番号

[0 0 0 0 0 0 2 9 5]

1 . 変更年月日

1 9 9 0 年 8 月 2 2 日

[変更理由]

新規登録

住 所

東京都港区虎ノ門 1 丁目 7 番 1 2 号

氏 名

沖電気工業株式会社